A quick reference guide to metagenome sequencing at JGI

June 2016 *Adam R. Rivers*

The DOE Joint Genome Institute is a user facility that provides sequencing and analysis to scientists from all over the world studying carbon cycling, bioenergy and biogeochemistry. These services are primarily provided through several competitive grant calls: an annual CSP call, a semiannual small-scale CSP call and joint Department of Energy user facility (FICUS) call. Each of these calls places caps on the total amount of sequencing that can be requested. The purpose of this guide is to describe the standard products the Metagenome Program offers and how to estimate the amount of sequencing needed.

We offer three standard product types: 1) metagenome sequencing, 2) metatranscriptome sequencing and 3) amplicon sequencing. Amplicon sequencing is designed to supplement metagenome and metatranscriptome sequencing projects so amplicon-only projects are not accepted. In addition to our standard products we have experimental sequencing products that we offer on a provisional basis. It is best to consult with JGI during the proposal writing process if you are interested in using experimental products. A list of the products we offer through the metagenome program is provided in Table 1. Many of these products are sequenced using Illumina HiSeq 2000 and 2500 1T sequencers. To help you in planning your sample allocations we have provided Table 2, which relates the gigabases of sequencing to read pairs and lane pooling estimates. The typical pooling degree used for

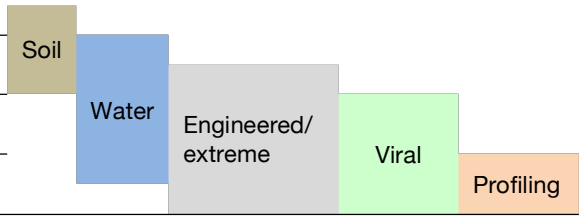
Table 1. The modes and typical yields of standard and experimental sequencing products from the Metagenome program.

	Product	Type (note)	Sequencing mode	Library creation (amount)	Typical sample pooling	Typical bases per sample	Typical sample Read pairs
Standard products	Metagenomes	Standard draft metagenome	Illumina HiSeq 2000/2500 1T 2x150 paired end, 270 bp insert	Truseq/ Kapa (500ng)	2-6 per lane	10-32GB	35-107 million
		Minimal draft metagenome			12-64 per lane	1- 5GB	3-17 million
	Meta-transcriptomes	rRNA depletion (by JGI)	Ilimuna HiSeq 2000/2500 1T 2x150 paired end, 270bp insert	Truseq/ Kapa (3µg-300ng)	2-6 per lane	10-32GB	35-107 million
		polyA selection (by JGI)					
	iTag amplicon sequencing	16S rRNA V4-V5 (plant PNA option)	Illumina MiSeq 2x300bp, up to 184 samples per run	PCR (150ng)	184 per run	81 MB	135,000
		18S rRNA V4					
		Fungal ITS2					
Experimental products	Improved draft metagenomes	Adds long reads to Illumina data (needs 5 µg high-quality DNA)	Pac Bio RSII 10Kb long libraries	SMRTbell	4 SMRT cells	600 MB	70,000 reads unpaired per SMRT cell
	RNA virus metagenomes		Illumina MiSeq 2x150 bp	Truseq/ (possibly Tigrt)	24	1GB	3 million
	DNA virus metagenomes	Double and single stranded genomes sequenced together	Illumina MiSeq, 2x150 bp Modified 300 bp shearing	Swift Bio.	24	1GB	3 million
	Stable Isotope probing (SIP) metagenomes	(We will sequence these but do not do the SIP processing)	Same as standard metagenome product				

some common environments is also listed in the table. The exact allocation of samples and degree of pooling is done collaboratively with JGI after proposals have been accepted. This guide should help you propose a reasonable sample allocation for you project based on our normal workflows. More detailed information on submitting samples is available on at JGI's sample submission page.

Table 2 Common pooling degrees, yields and environments sequenced with the HiSeq 2000 1T.

Degree of pooling	Bases (GB)	Read pairs (M)	Typical applications
1	64	212	
2	32	107	
3	21	71	
4	16	53	
6	11	36	
8	8	27	
12	5.3	14	
64 (uncommon)	1.0	3.3	



Frequently Asked Questions

Does JGI offer 2x250bp sequencing of metagenomes?
We do not currently offer 2x250bp sequencing. We have Illumina HiSeq 2500 machines and have tested long-read metagenome sequencing. However, Illumina's current sequencing reagents often result in poor read quality and a divergence in the GC content at the end of the reads that is not correctable with quality scores. We have provided Illumina with data on the problem and continue to test new regents. We are also considering alternatives such as running a shorter number of cycles for the second read.

Should I deplete my metatranscriptome samples?
We offer rRNA depletion as a standard service on metatranscriptome samples. We have achieved good success in many environments with our protocol, and users generally prefer to send total RNA.

Do you offer Nextera library amplification?
We do not currently offer Nextera for standard sequencing. In our evaluation it was difficult to control insert size and merged reads were often too short. Contamination was also high in our tests of low-concentration user samples, which resulted in high failure rates.